Osteoarthritis and Cartilage 19 (2011) 963–969

Osteoarthritis and Cartilage

OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

Definition of osteoarthritis on MRI: results of a Delphi exercise

D.J. Hunter †*, N. Arden ‡, P.G. Conaghan §, F. Eckstein ‖, G. Gold ¶#††, A. Grainger ‡‡, A. Guermazi §§, W. Harvey |||, G. Jones ¶¶, M.P. Hellio Le Graverand ##, J.D. Laredo †††, G. Lo ‡‡‡, E. Losina §§§ |||||, T.J. Mosher ¶¶¶, F. Roemer §§###, W. Zhang †††† on behalf of the OARSI OA Imaging Working Group

† *Rheumatology Department, Royal North Shore Hospital and Northern Clinical School, University of Sydney, Sydney, NSW, Australia*
‡ *NIHR Musculoskeletal Biomedical Research Unit, University of Oxford, Oxford, England, UK*
§ *Section of Musculoskeletal Disease, University of Leeds & NIHR Leeds Musculoskeletal Biomedical Research Unit, Leeds, UK*
‖ *Institute of Anatomy and Musculoskeletal Research, Paracelsus Medical University, Salzburg, Austria & Chondrometrics GmbH, Ainring, Germany*
¶ *Department of Radiology, Stanford University, Stanford, CA, USA*
# *Department of Bioengineering, Stanford University, Stanford, CA, USA*
†† *Department of Orthopaedic Surgery, Stanford University, Stanford, CA, USA*
‡‡ *Department of Radiology, Leeds Teaching Hospitals & NIHR Leeds Musculoskeletal Biomedical Research Unit, Leeds, UK*
§§ *Department of Radiology, Boston University School of Medicine, Boston, MA, USA*
||| *Division of Rheumatology, Tufts Medical Center, Boston, MA, USA*
¶¶ *Menzies Research Institute, Private Bag 23, Hobart, Tasmania, Australia*
Clinical Development and Medical Affairs, Pfizer, New London, CT, USA
††† *Department of Radiology, Hôpital Lariboisière, Assistance Publique des hôpitaux de Paris et Faculté de Médecine Paris 7, France*
‡‡‡ *Department of Medicine, Section of Immunology, Allergy and Rheumatology, Baylor College of Medicine, Houston, TX, USA*
§§§ *Department of Orthopedic Surgery, Brigham and Women's Hospital, Boston University School of Public Health, Boston, MA, USA*
||||| *Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA*
¶¶¶ *Department of Radiology, Penn State University College of Medicine, Hershey, PA, USA*
Department of Radiology, Klinikum Augsburg, Augsburg, Germany
†††† *Academic Rheumatology, The University of Nottingham, Nottingham, UK*

ARTICLE INFO

Article history:
Received 27 April 2010
Accepted 16 April 2011

Keywords:
Osteoarthritis
Magnetic resonance imaging
Diagnosis

SUMMARY

Objective: Despite a growing body of Magnetic Resonance Imaging (MRI) literature in osteoarthritis (OA), there is little uniformity in its diagnostic application. We envisage in the first instance the definition requiring further validation and testing in the research setting before considering implementation/feasibility testing in the clinical setting. The objective of our research was to develop an MRI definition of structural OA.

Methods: We undertook a multistage process consisting of a number of different steps. The intent was to develop testable definitions of OA (knee, hip and/or hand) on MRI. This was an evidence driven approach with results of a systematic review provided to the group prior to a Delphi exercise. Each participant of the steering group was allowed to submit independently up to five propositions related to key aspects in MRI diagnosis of knee OA. The steering group then participated in a Delphi exercise to reach consensus on which propositions we would recommend for a definition of structural OA on MRI. For each round of voting, $\geq$60% votes led to include and $\leq$20% votes led to exclude a proposition. After developing the proposition one of the definitions developed was tested for its validity against radiographic OA in an extant database.

Results: For the systematic review we identified 25 studies which met all of our inclusion criteria and contained relevant diagnostic measure and performance data. At the completion of the Delphi voting exercise 11 propositions were accepted for definition of structural OA on MRI. We assessed the diagnostic performance of the tibiofemoral MRI definition against a radiographic reference standard. The diagnostic performance for individual features was: osteophyte C statistic = 0.61, for cartilage loss C statistic = 0.73, for bone marrow lesions C statistic = 0.72 and for meniscus tear in any region C statistic = 0.78. The overall composite model for these four features was a C statistic = 0.59. We detected good specificity (1) but less optimal sensitivity (0.46) likely due to detection of disease earlier on MRI.

* Address correspondence and reprint requests to: D.J. Hunter, Rheumatology Department, Royal North Shore Hospital and Northern Clinical School, University of Sydney, Sydney, NSW 2066, Australia. Tel: 61-2-9926-7379; Fax: 61-2-9906-1859.
E-mail address: david.hunter@sydney.edu.au (D.J. Hunter).

1063-4584/$ – see front matter
doi:10.1016/j.joca.2011.04.017

Conclusion: We have developed MRI definition of knee OA that requires further formal testing with regards their diagnostic performance (especially in datasets of persons with early disease), before they are more widely used. Our current analysis suggests that further testing should focus on comparisons other than the radiograph, that may capture later stage disease and thus nullify the potential for detecting early disease that MRI may afford. The propositions are not to detract from, nor to discourage the use of traditional means of diagnosing OA.

Introduction

The American College of Rheumatology (ACR) classification criteria have been the mainstay for diagnosing osteoarthritis (OA) both in the clinic and in research settings[1]. The clinical criteria include a combination of the patient's age, signs and symptoms on physical exam, and further criteria also include the addition of radiographic and/or laboratory evidence. When the radiograph is used along with a physical exam, the sensitivity and specificity of this combined method for OA diagnosis are 91% and 86%, respectively. The standard against which the classification criteria were judged was the clinical diagnosis of OA. This radiograph based diagnostic method is cheap and readily available. Although the ACR combined clinical and radiographic criteria remains the standard for diagnosis both in the clinic and in research, it relies upon conventional radiographs which means they may be less sensitive to detecting early structural changes of OA.

In a more recent effort to define OA, the European League Against Rheumatism (EULAR) OA Task Force suggested that a confident clinical diagnosis of knee OA may be made according to three symptoms (persistent knee pain, morning stiffness and reduced function) and three signs (crepitus, restricted movement and bony enlargement)[2]. Part of the research agenda posed by this group was to examine the diagnostic relevance of MRI for OA.

MRI potentially affords inherent advantages for a structural diagnosis of OA of whole organ assessment with multiplanar acquisitions. The potential for MRI to be more sensitive to earlier disease, detecting change, and the capacity of this technology to visualize joint structural changes beyond gross changes in bone and in the joint space, has resulted in great interest in the use of MRI for assessing diagnostic status, disease severity and monitoring progression[3,4]. Increasingly, MRI is being used in clinical practice to facilitate diagnostic decisions. Because of cost concerns, as well as lack of clarity about diagnostic performance and little standardization regarding MRI interpretation, it is unclear, however, whether this increased use of MRI in clinical practice is rational. In the research setting, some have proposed recruiting persons with early disease as defined on MRI into Disease Modifying Osteoarthritis Drug (DMOAD) trials where currently no existing definition exists. Due to these inherent concerns about the application of MRI to facilitate OA diagnosis it is important that this be more rigorously tested. Despite a growing body of MRI literature in OA, there is little uniformity in its diagnostic application. This lack of uniformity may stem from the absence of criteria for an MRI OA structural diagnosis. We would envisage in the first instance the development of an MRI definition would require further validation and testing in the research setting before considering implementation/feasibility testing in the clinical setting.

In the absence of MRI criteria for OA diagnosis there are a number of different methods that could be used to reach consensus. One of the methods used more commonly in the medical literature is the Delphi process[5,6]. With this background, the objective of our research was to develop an MRI definition of structural OA. More specifically, the intent was to find what structural changes on MRI would constitute a structural diagnosis of OA and develop testable definitions of OA (knee, hip and/or hand) on MRI. In addition, we began the formal assessment of the diagnostic performance of the proposed definition although this will require much further testing.

Materials and methods

The MRI-based definition of structural OA was developed using an evidence driven approach with results of a systematic review provided to an expert group **prior** to a Delphi exercise. We undertook a multistage process consisting of a number of steps.

1. *Systematic Review*. As part of a larger Osteoarthritis Research Society International (OARSI) Food and Drug Administration (FDA) OA initiative we undertook a systematic review of the OA MRI literature for the purpose of assessing the clinometric properties of MRI measurements in knee OA. An online literature search was conducted of the OVID MEDLINE (1945-), EMBASE (1980-) and Cochrane databases (1998-) of articles published up to the time of the search, April 2009, with the search entries "MRI", and "osteoarthritis". All articles which used MRI, in some form, on patients with osteoarthritis of the knee, hip, or hand were included. Only studies published in English were included. Studies presenting non-original data were excluded, such as reviews, editorials, opinion papers, or letters to the editor. In addition to the clinometric properties (reliability, responsiveness and validity are the focus of separate ongoing analyses) of MRI in OA, details on rigor of study design to construct the Downs methodological quality score[7], we also extracted data pertaining to the diagnostic performance of MRI. Data was extracted on the diagnostic performance of MRI or material pertaining to defining OA on MRI which is the focus of this manuscript. The ability of MRI to discriminate between patients with and without knee OA (diagnosed using different reference standards including radiography and arthroscopy) was summarized by sensitivity, specificity, positive and negative predictive values (PPV, NPV). The pooled data from the literature search were shared with the OA Imaging Working Group *via* an interactive website and other interested members of the OA community during a public meeting.
2. *Steering group*. A multidisciplinary, geographically diverse steering group was selected on the basis of content and/or methods expertise and was invited to participate in the OA MRI definition development project by OARSI. The steering group (the listed co-authors) met by teleconference and were invited to propose MRI definitions of OA. Each participant was allowed to submit independently up to five propositions related to key aspects in MRI diagnosis of knee OA.
3. *Delphi exercise*. The steering group then participated in a Delphi exercise to reach consensus on which propositions we would recommend for a definition of structural OA on MRI[5]. For each round of voting, $\geq$60% votes led to include and $\leq$20% votes led to exclude a proposition. Propositions receiving 20–60% votes were discussed and another round of voting was taken. The

process was repeated until all propositions were either included or excluded.

4. *Scientific Community Feedback*. The systematic review data on the diagnostic performance of MRI in OA and the MRI OA definition propositions after initial rounds of the Delphi exercise were presented at a public meeting in Montreal (September 2009) during the OARSI annual scientific meeting to gain feedback on the initiative. The scientific community feedback occurred between rounds 2 and 3 of the Delphi exercise and was used to inform wording for the Delphi propositions for the third round.
5. *Diagnostic performance in Osteoarthritis Initiative (OAI)*. To move beyond consensus and to test the validity or otherwise of the definition requires assessing the diagnostic performance of the new definition against the current reference standard. There are a number of means of defining OA but the most commonly used means in current research studies is the use of the Kellgren and Lawrence (KL) grade, with those having OA defined as $\geq$2. To this end we have assessed the MRI definition of tibiofemoral OA (definition 10) as proposed in this manuscript against this reference standard. We also acknowledge that other constructs could be used for comparison and that further assessing the validity of the proposed definitions is required. This analysis was conducted in an existing dataset that we have previously published[8]. These are 160 participants from the progression sub-cohort of the OAI Study, an ongoing multi-center study, focusing on knee OA. Sixteen percent of the study sample did not have radiographic OA using the commonly accepted criteria of KL grade $\geq$ 2. For this analysis we have included osteophyte, cartilage loss, bone marrow lesion, and meniscus. Statistical analyses were performed using SAS software, (SAS Institute Inc, Cary, North Carolina, release 8.2). This analysis was conducted after development of the definition. Further given the timing of conducting these analyses, after the Delphi was complete, they did not inform that process.

Results

Systematic literature review

We identified 25 studies which met all of our inclusion criteria and contained relevant diagnostic measure and performance data. The studies included a total of 2704 participants, 1733 with OA and 971 without OA. Three of the studies focused on the hip, one on the hand and the remainder focused upon the knee. A large majority of the studies used semi-quantitative MRI measurement techniques for measuring OA. The only other measurement technique appearing more than once was quantitative measurement of cartilage morphology (thickness, volume). Various tissues were imaged in the different studies. In the majority of studies cartilage was examined (19 studies). The two other commonly viewed tissue types were meniscus (eight studies) and synovium (three studies). As the gold standard measure against which the MRI diagnostic techniques were compared, arthroscopy was used most frequently, followed by radiographs, and histological section. Table I reflects the diagnostic performance of all MRI techniques against the various gold standards. The diagnostic performance of MRI varied markedly depending on which reference standard it was compared to. In general the MRI performance was better when it was compared to the tissue of interest directly (e.g., arthroscopy, histology) as distinct from X-ray.

Delphi results

The steering group members provided a total of 53 propositions. Prior to the first round of voting, any redundant (overlapping) propositions were amalgamated. During the first round of voting, five propositions were rejected and one was accepted.

The second round commenced with 19 propositions for voting. During the second round a further two propositions were rejected and a further seven were accepted and one redundant proposition was combined.

The third round commenced with nine propositions requiring further voting. After completion of the third round, three of these were accepted and a further six were rejected. Thus at the completion of the Delphi voting exercise 11 propositions were accepted for definition of structural OA on MRI (Table II).

The first nine of the propositions listed in the table provide an important conceptual framework or "preamble" for the MRI definition of structural OA. Propositions 10 and 11 are testable definitions in other datasets.

Table I
Diagnostic performance of all MRI techniques measured against gold standards, separated by gold standard used

Gold standard used		Mean	Median	Min	Max	No. of entries
Arthroscopy (23 studies)	Sensitivity	68.7%	72%	17%	100%	66
	Specificity	88%	93.9%	58%	99.96%	44
	PPV	75%	84.7%	48%	92.2%	7
	NPV	92.2%	94%	78%	100%	7
X-ray (20 studies)	Sensitivity	56.4%	60%	18%	90%	11
	Specificity	85.2%	85%	67%	96%	11
	PPV	49%	49%	48%	50%	2
	NPV	87%	87%	86%	88%	2
Histology (17 studies)	Sensitivity	68.4%	69%	36%	93%	17
	Specificity	75.5%	72.4%	62%	100%	5

Assessing diagnostic performance of proposed definition

The diagnostic performance of the proposed definition was tested in 160 participants from the progression sub-cohort of the OAI Study[8]. Sixteen percent of the study sample did not have radiographic OA using the commonly accepted criteria of KL grade $\geq$ 2.

Osteophyte

The analysis of the osteophyte volume using the Dual Echo Steady State (DESS) sequences was done within the segmentation process from a former study[8], investigating cartilage morphometry changes in participants out of the OAI database. For this analysis osteophyte volume in the tibiofemoral joint was assessed as the sum of osteophyte volume at the medial and lateral tibia as well as medial and lateral femur. The reference standard for comparison of diagnostic performance was radiographic OA; defined as KL grade ≥ 2[9] in the same knee.

For osteophyte volume:

(a) cutpoint = 50 mm^3, c = 0.53
(b) cutpoint = 100 mm^3, c = 0.52
(c) cutpoint = 200 mm^3, c = 0.61 [see Fig. 1(a)].

Full thickness cartilage loss

The cartilage segmentation was done using DESS MRI sagittal sequences acquired by the OAI. The DESS sequence provides a complete high-resolution view of the knee cartilage tissue with good contrast and separation between fluid, cartilage, meniscus and bony tissue. The segmentation, analysis methods and proprietary software have been previously described[8,10,11]. After image segmentation, the following measures were analyzed:

Table II
Accepted propositions for definition of OA on MRI after Delphi voting completion

Preamble	
1	MRI changes of OA may occur in the absence of radiographic findings of OA.
2	MRI may add to the diagnosis of OA and should be incorporated into the ACR diagnostic criteria including X-ray, clinical and laboratory parameters.
3	MRI may be used for inclusion in clinical studies according to criteria defined above but should not be a primary diagnostic tool in a clinical setting.
4	Certain MRI changes in isolation including cartilage loss, cartilage compositional change, cystic change and bone marrow lesions, ligamentous and tendinous damage, meniscal damage, and effusion and synovitis are not diagnostic of osteoarthritis.
5	No single MR finding is diagnostic of knee OA.
6	MRI findings indicative of knee OA may include abnormalities in all tissues of the joint: bone, cartilage meniscus, synovium, ligament and capsule.
7	Given the multiple tissue abnormalities detected by MRI in OA, diagnostic criteria are likely to involve several possible combinations of features.
8	Definite osteophyte formation is indicative of osteoarthritis.*
9	Joint space narrowing assessed by (non-weight bearing) MRI cannot be used as a diagnostic criterion.
Definitions	
10	A definition of tibiofemoral osteoarthritis on MRI would be: The presence of both group [A] features or one group [A] feature and two or more group [B] features Group [A] after exclusion of joint trauma within the last 6 months (by history) and exclusion of inflammatory arthritis (by radiographs, history and laboratory parameters): i) Definite osteophyte formation* ii) Full thickness cartilage loss Group [B]: i) Subchondral bone marrow lesion or cyst not associated with meniscal or ligamentous attachments ii) Meniscal subluxation, maceration or degenerative (horizontal) tear iii) Partial thickness cartilage loss (where full thickness loss is not present) iv) Bone attrition
11	Definition of PF OA requires all of the following involving the patella and/or anterior femur: i) A definite osteophyte ii) Partial or full thickness cartilage loss

* The definition of a 'definite osteophyte' was not delineated in the Delphi process and requires further validation.

1. Cartilage volume.
2. Normalized Cartilage volume (Volume normalized to bone surface interface area).
3. Denuded area (Total Cartilage Bone Interface area denuded of cartilage). The denuded area is the area of bone where a full thickness cartilage defect is present.

For cartilage loss we used the denuded area and depicted this at two different cutpoints based upon prior data[12]:

(a) cutpoint = 0 mm^2, c = 0.65
(b) cutpoint = 10 mm^2, c = 0.73 [see Fig. 1(b)].

Bone marrow lesion

On the sagittal intermediate-weighted (IW) Turbo Spin Echo (TSE) fat-suppressed images (FS) a Bone Marrow Lesion (BML) was described when seeing an irregular hyperintense signal in the subchondral bone, proximal to the epiphyseal line. The size of the BMLs was evaluated for size from 0 to 3 at each of the following locations using Boston–Leeds Osteoarthritis Knee Score (BLOKS)[13]: medial and lateral weight-bearing femur, and medial and lateral tibia. We classified BMLs as present in any region (>0) and also those with BLOKS score > 1 in any region.

For BML:

(a) cutpoint = 0, c = 0.54
(b) cutpoint = 1, c = 0.72 [see Fig. 1(c)].

Degenerative meniscal tear

The sagittal IW TSE FS sequences, were used to score the meniscal integrity using the Whole-Organ Magnetic Resonance Imaging Score (WORMS) grading system[14]. The anterior horn, body segment and posterior horn of each of the medial and lateral menisci were graded from 0 to 4 based on both the sagittal and coronal images: 0 = intact; 1 = minor radial tear or parrot-beak tear; 2 = nondisplaced tear; 3 = displaced tear or partial resection; 4 = complete maceration/destruction or complete resection.

For meniscal abnormality we scored this as present in any region (>0) and also those with score ≥ 2.

For meniscus:

(a) cutpoint = 0, c = 0.5
(b) cutpoint = 2, c = 0.78 [see Fig. 1(d)].

Composite

We also assessed the formation of a composite model including all of these features. For the composite model we used the most discriminatory cutpoints from the above features. Thus for osteophyte we used a threshold of 200 mm^3 (C statistic for individual model = 0.61), for cartilage loss we used a cutpoint of 10 mm^2 (C statistic for individual model = 0.73), for BML > 1 (C statistic for individual model = 0.72) and for meniscus ≥ 2 in any region (C statistic for individual model = 0.78). Overall composite model C statistic = 0.59 [see Fig. 1(e)].

Discussion

OA is a complex disease characterized by involvement of multiple tissues in the synovial joint[15]. The results of the Delphi process present a number of propositions that are statements of preamble and context setting (propositions one to nine). By their nature the preamble statements will be difficult to formally test and assist here in providing context for the two definitions proposed.

For clarification the statement in proposition eight may appear to contradict that in proposition five however the use of the word "indicative" which ultimately means suggestive, means that it is not incontrovertible proof of OA.

The impetus behind imaging developments in OA is primarily to facilitate therapeutic development of interventions that modify joint structure. A number of concerns have been raised about the feasibility of disease modifying approaches using current plain radiographic technologies, as this captures persons with late stage disease that may not be amenable to such interventions[16]. The motivation to include MRI as a means of defining disease is with the intent that one may be able to identify early, pre-radiographic disease, thus enabling recruitment of study populations where structure modification (or structure maintenance) may be realistic in a more preventative manner. Prior to using the definitions that have been developed, it is important that they be adequately tested.

We have begun this process but it requires much more assessment particularly focusing on cohorts with early, pre-radiographic disease. The overall composite model C statistic was a disappointing 0.59. It is important to indicate that there was a small number of persons without radiographic OA in this dataset and the specificity for this diagnosis was =1. Not surprisingly given the potential for MRI to detect earlier disease the sensitivity was 0.46.

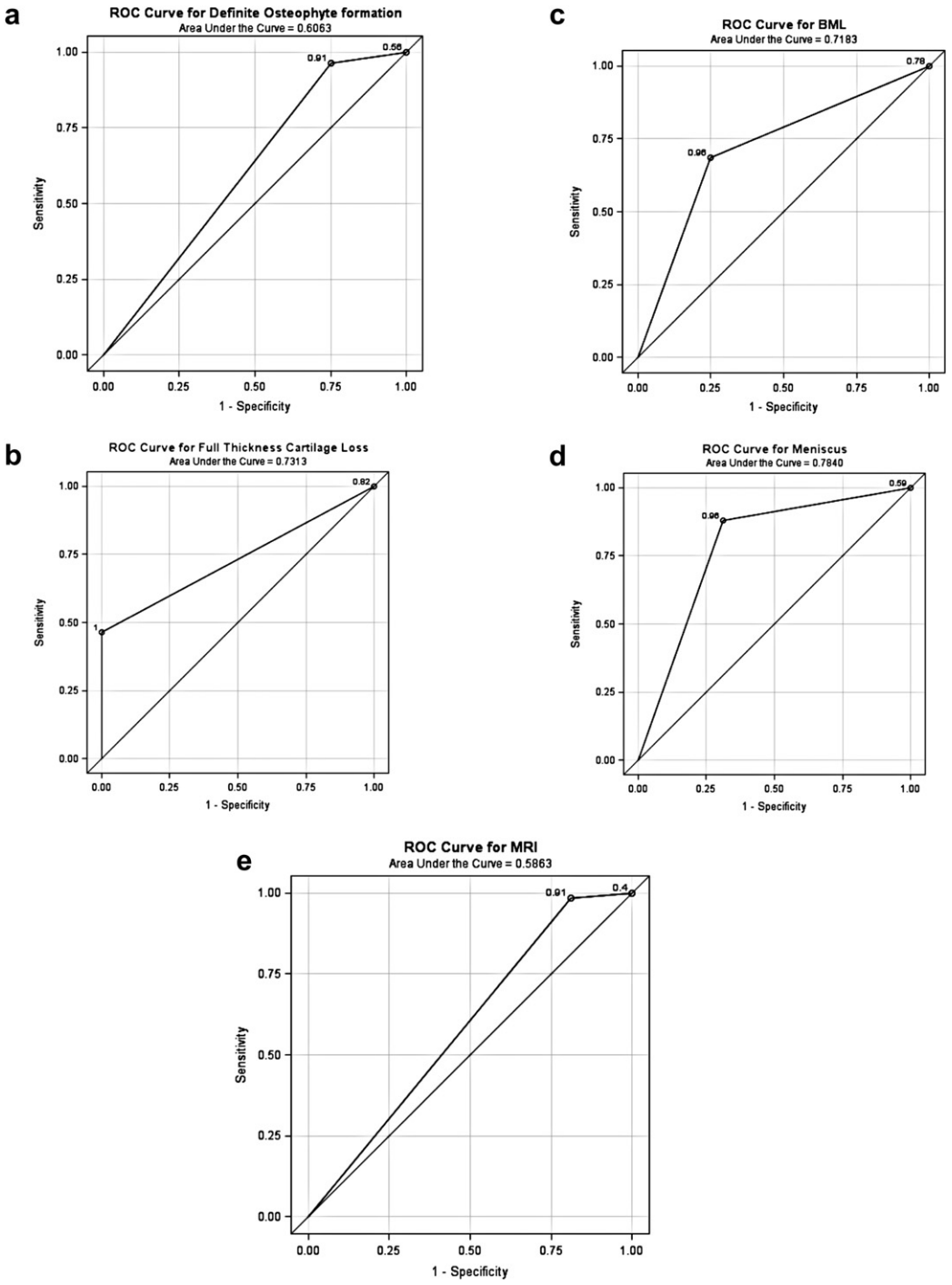

Fig. 1. (a) Osteophyte volume – the Receiver Operating Characteristic (ROC) curve for the cutpoint of 200 mm^3. (b) Cartilage loss – the ROC curve for the cutpoint of 10 mm^2. (c) BML – the ROC curve for the cutpoint of BML > 1 in any region. (d) Meniscus – the ROC curve for the cutpoint of meniscal score ≥ 2 in any region. (e) Composite ROC curve.

Further assessment of these definitions is required in larger datasets both using the construct of radiographic OA as well as other endpoints including the presence of symptoms.

The next step in testing the validity of the definitions developed would be to formally test their diagnostic performance against other current diagnostic standards, ideally in at least two independent datasets. Many of the propositions do not lend themselves to formal validity testing, being statements to create context or to set limitations of MRI in this arena. Propositions 10 and 11 do, however, offer an opportunity for formal testing against other diagnostic constructs, such as plain radiography or clinical diagnosis.

The lack of a universally agreed reference standard does, however, create a challenge with regards performing such analyses. MRI has the potential to diagnose OA earlier than the current reference standard of radiography and there is little consensus on what represents OA using other potential reference standards such as arthroscopy. The lack of a recognized gold standard and the variety of gold standards used in the systematic review prior to the Delphi exercise may have introduced complexity that could have adversely influenced decision making in the Delphi voting. Novel methods may assist here in the absence of a perfect gold standard that make estimates of the true prevalence based on the outcome of different tests[17,18].

The goal of this exercise was to bring leading experts in OA with some expertise in MRI together to agree as much as possible on definitions that could be published and suggested for free use in future studies. Ultimately, however, the exercise was primarily dependent upon an expert consensus based approach[19]. This has limitations that need to be acknowledged including that it is based on the subjective opinion of the participants, and the questionable premise that 'pooled intelligence' enhances individual judgement and captures the collective opinion of experts[6].

It is important to recognize that the propositions have been developed for structural OA, not for a clinical diagnosis, not for early OA, and not to facilitate staging of the disease. Some of the elements of different propositions have questionable clinical relevance including the osteophyte that may simply be an adaptive/reparative response to altered alignment[20]. These further steps in developing the utility of MRI in the setting of OA will naturally follow from this work.

In addition to this work, the OARSI OA Imaging Working group is intending to develop a core set of sequences that can facilitate application of MRI in OA for diagnosis, staging and assessing progression. Some modification of existing scoring techniques may be required to facilitate these goals[14,15].

Using a modified Delphi approach we have developed 11 propositions for definition of OA on MRI. The aim for their development is that they be formally tested regarding their diagnostic performance, before they are more widely used. The propositions are not to detract from, nor to discourage the use of traditional means of diagnosing OA.

In our preliminary analysis of the diagnostic performance of the tibiofemoral definition we detected good specificity but less optimal sensitivity that is likely due to detection of disease earlier on MRI. Our current analysis suggests that further testing should focus on comparisons other than the radiograph, that may capture later stage disease and thus nullify the potential for detecting early disease that MRI may afford.

Conflict of interest

The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

David Hunter receives research or institutional support from AstraZeneca, DonJoy, Lilly, Merck, NIH, Pfizer, Stryker and Wyeth.

Felix Eckstein is Chief Executive Officer (CEO) of Chondrometrics GmbH, a company providing MR image analysis services. He provides consulting services to MerckSerono and Novartis, has received speaking honoraria from Medtronic, Synthes, Merck, Genzyme, Glaxo Smith Kline, and Sanofi Aventis, and has received research support from Pfizer, Eli Lilly, MerckSerono, Glaxo Smith Kline, Centocor, and Novartis.

Garry Gold receives research support from GE Healthcare and is a consultant for Zimmer, Arthrocare, and ICON Medical.

Ali Guermazi receives research or institutional support from NIH and GE Healthcare. He is consultant to MerckSerono, Stryker, Genzyme and Novartis. He is President of BICL (Boston Imaging Core Lab), Limited Liability Company (LLC), a company providing radiologic image assessment services.

Marie Pierre Hellio Le Graverand is a full time employee of Pfizer Inc.

Jean Denis Laredo receives research support from Servier and Pierre Fabre.

Tim Mosher serves as a paid consultant for Kensey Nash Corporation.

Frank Roemer is shareholder of Boston Imaging Core Lab (BICL) LLC, a company providing radiologic image assessment services.

Weiya Zhang receives research support from The Arthritis Research UK, The BUPA Foundation, EULAR, OARSI, Savient.

Other authors declared no conflict of interest.

Contributions

DJH conceived and designed the study, drafted the manuscript and takes responsibility for the integrity of the work as a whole, from inception to finished article. EL and WZ were also involved in the design of the study. All authors contributed to acquisition of the data. All authors critically revised the manuscript and gave final approval of the article for submission.

Acknowledgements

We recognize the invaluable support of Valorie Thompson for administrative and editorial support, the OA community for their feedback and OARSI for their invaluable support of this activity. We would also like to recognize the invaluable support of Kelly Hirko and Leo Menashe who assisted with the systematic review.

References

1. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, *et al*. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Arthritis Rheum 1986;29: 1039–49.
2. Zhang W, Doherty M, Peat G, Bierma-Zeinstra MA, Arden NK, Bresnihan B, *et al*. EULAR evidence-based recommendations for the diagnosis of knee osteoarthritis. Ann Rheum Dis 2010; 69:483–9.
3. Guermazi A, Burstein D, Conaghan P, Eckstein F, Hellio Le Graverand-Gastineau MP, Keen H, *et al*. Imaging in osteoarthritis. Rheum Dis Clin North Am 2008;34:645–87 [Review] [183 refs].
4. Hunter D, Hellio Le Graverand M, Eckstein F. radiologic markers of osteoarthritis progression. Curr Opin Rheumatol 2009;21:110–7.
5. Jones J, Hunter D. Consensus methods for medical and health services research. BMJ 1995;311:376–80.
6. Mullen PM. Delphi: myths and reality. J Health Organ Manag 2003;17:37–52.
7. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998;52:377–84.
8. Hunter DJ, Niu J, Zhang Y, Totterman S, Tamez J, Dabrowski C, *et al*. Change in cartilage morphometry: a sample of the progression cohort of the Osteoarthritis Initiative. Ann Rheum Dis 2009;68:349–56.
9. Kellgren JH, Lawrence JS. Atlas of Standard Radiographs. Oxford: Blackwell Scientific; 1963.
10. Tamez-Pena J, Barbu-McInnis M, Totterman S. Knee cartilage extraction and bone-cartilage interface analysis from 3D MRI data sets. Proc SPIE Med Imag 2004.
11. Tamez-Pena J, Barbu-McInnis M, Lerner A, Totterman S. Unsupervised definition of the tibia-femoral joint regions of the human knee and its applications to cartilage analysis. Proc SPIE Med Imag 2006.
12. Hunter D, Niu J, Zhang Y, LaValley M, McLennan C, Hudelmaier M, *et al*. Premorbid knee OA is not characterized by diffuse thinness: the Framingham Study. Ann Rheum Dis 2008;67:1545–9.
13. Hunter DJ, Lo GH, Gale D, Grainger AJ, Guermazi A, Conaghan PG. The reliability of a new scoring system for knee osteoarthritis MRI and the validity of bone marrow lesion assessment: BLOKS (Boston Leeds Osteoarthritis Knee Score). Ann Rheum Dis 2008;67:206–11.

14. Peterfy CG, Guermazi A, Zaim S, Tirman PF, Miaux Y, White D, *et al*. Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis. Osteoarthritis Cartilage 2004;12:177–90.
15. Nuki G. Osteoarthritis: a problem of joint failure. Zeitschrift Rheumatol 1999;58:142–7 [Review] [55 refs].
16. Hunter DJ, Hellio Le Graverand-Gastineau MP. How close are we to having structure-modifying drugs available? Rheum Dis Clin North Am 2008;34:789–802 [Review] [79 refs].
17. Hui SL, Walter SD. Estimating the error rates of diagnostic tests. Biometrics 1980;36:167–71.
18. Enoe C, Georgiadis MP, Johnson WO. Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. Prev Vet Med 2000;45:61–81.
19. Sniderman AD. Clinical trials, consensus conferences, and clinical practice. Lancet 1999;354:327–30.
20. Felson DT, Gale DR, Elon GM, Niu J, Hunter DJ, Goggins J, *et al*. Osteophytes and progression of knee osteoarthritis. Rheumatology (Oxford) 2005;44:100–4.